HHS Public Access

Author manuscript

J Biopharm Stat. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

J Biopharm Stat. 2016 ; 26(3): 507–518. doi:10.1080/10543406.2015.1052480.

ESTIMATION OF TREATMENT EFFECT IN A SUB-POPULATION: AN EMPIRICAL BAYES APPROACH

Changyu Shen[1], **Xiaochun Li**[1], and **Jaesik Jeong**[2]

[1]Department of Biostatistics, School of Medicine, Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA

[2]Department of Statistics, Chonnam National University, Gwangju, Korea

Abstract

It is well recognized that the benefit of a medical intervention may not be distributed evenly in the target population due to patient heterogeneity and conclusions based on conventional randomized clinical trials may not apply to every person. Given the increasing cost of randomized trials and difficulties in recruiting patients, there is a strong need to develop analytical approaches to estimate treatment effect in sub-populations. In particular, due to limited sample size for sub-populations and the need for multiple comparisons, standard analysis tends to yield wide confidence intervals of the treatment effect that are often non-informative. We propose an empirical Bayes approach to combine both information embedded in a target sub-population and information from other subjects to construct confidence intervals of the treatment effect. The method is appealing in its simplicity and tangibility in characterizing the uncertainty about the true treatment effect. Simulation studies and a real data analysis are presented.

Keywords

Causal inference; Empirical Bayes; Heterogeneity in treatment effect; Sub-group analysis

1. INTRODUCTION

The primary goal of typical randomized clinical trials is the assessment of the effect of a medical intervention as compared with an appropriate control or reference intervention. A well-accepted principle is the characterization of the clinical benefit of the intervention by the average treatment effect (ATE), which is the difference in the expectation of the outcome over the entire population under control and intervention. Nevertheless, it is well-known that patient heterogeneity may lead to heterogeneity in the treatment effect (e.g. the intervention has different impact on different people) (Davidoff, 2009; Kent and Hayward, 2007). Therefore, in many clinical trials, pre-specified, well-defined sub-populations are examined separately to study heterogeneity in treatment effect (Wang et al., 2007). There are two main limitations in this type of sub-group analysis. First, tests of treatment effects in sub-populations tend to be under-powered due to smaller sample sizes and more stringent type I

Address correspondence to Changyu Shen, Department of Biostatistics, School of Medicine, Fairbanks School of Public Health, Indiana University, 410 West 10th Street, Suite 3000, Indianapolis, IN 46202; chashen@iu.edu.

error control if multiple comparisons are made. Second, the pre-specified sub-populations may not coincide with the ones with large or small ATEs, and thus the analysis is targeted on the wrong groups of patients. As drug development has become increasingly expensive with high failure rates, there is a strong motivation to explore sub-populations with large treatment effects in a *post hoc* manner. The identification of a sub-population for which the intervention is effective will benefit patients, and is of great interest to both the pharmaceutical industry and the regulatory agencies. Towards this goal, several statistical methods have been proposed, which are based on either optimization of the expected outcome over the space of regimens (Qian and Murphy, 2011; Zhao et al., 2012), clustering patients by a data-driven score followed by estimation of treatment effect of each cluster (Cai et al., 2011; Zhao et al., 2013), classification tree based methods that divide the entire population into clusters with different treatment effects (Foster, Taylor and Ruberg, 2011; Lipkovich et al., 2011), and full Bayesian approach (Berger, Wang and Shen, 2014). These approaches offer tools to search for sub-populations with distinct treatment effects.

In this article, we focus on the first issue raised in previous paragraph, that is, the decreased precision in estimating treatment effect in a pre-specified sub-population due to limited sample size and possible multiple comparison adjustment. Thus, we are not primarily concerned with the potentially strong bias induced by *post hoc* selection of sub-populations with large treatment effect. One solution to the precision loss is to borrow information from other subjects using regression models, which has been well recognized and adopted in practice. We propose an empirical Bayes (EB) approach to construct confidence intervals for the treatment effect in a sub-population for a binary outcome. The EB is a well-established method to estimate parameter of one unit by borrowing information from other units (Efron, 1996, 2010a; Morris, 1983). Central to our approach is the conceptualization and estimation of a prior distribution for the treatment effect in the sub-population. The EB approach will treat the treatment effect estimate based on data from the given sub-population as the “direct evidence”, and the prior distribution estimated from entire data as the “indirect evidence” (Efron, 2010a). Thus, we borrow information from other people through construction and estimation of the prior. It represents a compromise between two “extreme” approaches. On one end of the spectrum, the inference is solely based on the data of the given sub-population and data from other people are deemed “irrelevant”. On the other end, the treatment effect is assumed to be the same for all sub-populations and the entire data are used to infer the common treatment effect. The EB offers a natural way to combine both extremes (Efron, 1996). Closely related approaches are the full Bayesian approach in the setting of linear models (Dixon and Simon, 1991; Jones et al., 2011) and EB approach assuming a normal prior (Davis and Leffingwell, 1990; Louis, 1984). To our knowledge, this is the first attempt of applying EB approach without the need to assume a parametric family as the prior distribution in the setting of treatment effect estimation in sub-populations. The advantages of our method are three folds. First, the posterior distribution has an appealingly natural and intuitive interpretation as explained in Section 2. Second, by using the data to estimate the prior, we maintain objectivity in the analysis. Third, the procedure can be fully automated for any disease and intervention without the need to adapt the specification of the prior to various diseases and interventions. Fourth, the posterior distribution of the treatment

effect offers a convenient tool to estimate false discovery rate (FDR) when multiple sub-populations are evaluated (Benjamini and Hochberg, 1995; Efron, 2010b).

In what follows, we will describe our method in Section 2, apply it to the MAGnesium In Coronaries (MAGIC) trial in Section 3, present a simulation study in Section 4 and conclude the article with a discussion in Section 5.

2. METHOD

2.1 Description of the Problem

We consider two-arm randomized clinical studies targeted on some patient population with a binary outcome. A group of subjects who meet certain criteria defined by baseline characteristics will be called a *sub-population*. For instance, the population can be all adults with type 2 diabetes mellitus, and an example sub-population is composed of those who are female, age between 40 and 50 years, and currently taking oral medications. We focus on using discrete baseline characteristics to define sub-populations, but the method can be easily extended to continuous covariates (see Section 5). Specifically, suppose there are k characteristics, c_j, $j = 1, 2,\ldots, k$, each with L_j levels (e.g. $L_j = 2$ if c_j is binary). In theory, there are in total $\prod_{j=1}^{k} L_j$ "cells", or smallest sub-populations that cannot be further divided using the k characteristics. In a realistic data set, many of these cells are empty, and the actual number of cells with at least one unit, L, is smaller. The L non-empty cells can yield $S = 2^L - 2$ non-empty sub-populations (not including the entire population). This number is rather large, which offers a great opportunity to conceptualize a prior distribution. For example, with eight binary variables, there are in theory 256 cells. Even if only 20% is non-empty, there are still $2^{51} - 2 \approx 10^{15}$ sub-populations.

2.2 The Prior Distribution

We will use a binary vector $\mathbf{Z} = (Z_1, Z_2,\ldots, Z_L)$ to label each sub-population, where Z_j=1 means the jth cell is included in the sub-population and 0 otherwise (j=1,2,…, L). For each $\mathbf{Z}$, there are three parameters that can be estimated: the proportion of the population that falls in the sub-population or the *size* of the sub-population ($\theta_1(\mathbf{Z})$), the event rate in the control arm ($\theta_2(\mathbf{Z})$), and the event rate in the intervention arm ($\theta_3(\mathbf{Z})$). Let $\boldsymbol{\theta}(\mathbf{Z}) = (\theta_1(\mathbf{Z}), \theta_2(\mathbf{Z}), \theta_3(\mathbf{Z}))^T$. At the conceptual level, the collection of the S $\boldsymbol{\theta}(\mathbf{Z})$ values induce a distribution on the space of $(0,1)^3$, which is a natural choice of the prior distribution. It can be viewed as infinite past "experience" (Efron, 2012), where "experience" in this case refers to the true value of $\boldsymbol{\theta}$ for each of the S sub-populations. Such a distribution can also be viewed as induced by treating $\boldsymbol{\theta}(\mathbf{Z})$ as a random vector, where each component Z_j is independently and identically distributed as a Bernoulli variable with probability of success equal to p=0.5. In fact, we can set p to different values so that the prior puts more weight on those sub-populations with close to $[Lp]$ cells ([.] is the rounding operation). For instance, if L=100 and the sub-population of interest has 30 cells, then it seems natural to use p=0.3 for the definition of the prior since sub-populations with 30 cells are more "relevant" to the given sub-population.

To characterize the prior distribution, we introduce some notations. Let τ_j, α_j, and β_j be the true values of the size, the event rate in the control arm, and the event rate in the intervention arm for cell j (j=1,2,…,L). In the Appendix, we show that for large L, the prior distribution depends on $\boldsymbol{\lambda} = (\boldsymbol{\mu},\boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu}=(\mu_1,\mu_2,\mu_3)^T=p\left(1,\sum_{j=1}^{L}\tau_j\alpha_j,\sum_{j=1}^{L}\tau_j\beta_j\right)^T$$

$$\sum=\begin{pmatrix}\sum_{\mathbf{11}} & \sum_{\mathbf{12}}\\ \sum_{\mathbf{21}} & \sum_{\mathbf{22}}\end{pmatrix},\ \sum_{\mathbf{11}}=p(1-p)\sum_{j=1}^{L}\tau_j^2,\ \sum_{\mathbf{12}}=\sum_{\mathbf{21}}^{T}=p(1-p)\left(\sum_{j=1}^{L}\tau_j^2\alpha_j,\sum_{j=1}^{L}\tau_j^2\beta_j\right),$$

$$\sum_{\mathbf{22}}=p(1-p)\begin{pmatrix}\sum_{j=1}^{L}\tau_j^2\alpha_j^2 & \sum_{j=1}^{L}\tau_j^2\alpha_j\beta_j\\ \sum_{j=1}^{L}\tau_j^2\alpha_j\beta_j & \sum_{j=1}^{L}\tau_j^2\beta_j^2\end{pmatrix}.$$

Specifically, due to the Lindeberg-Feller central limit theorem, the prior distribution can be approximated by

$$p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})=p_{\boldsymbol{\lambda}}(\theta_1)p_{\boldsymbol{\lambda}}(\theta_2,\theta_3|\theta_1)$$
$$=N(\theta_1;\mu_1,\sum_{\mathbf{11}})\times N\left((\theta_2,\theta_3)^T;[(\mu_2,\mu_3)^T+\sum_{\mathbf{21}}\sum_{\mathbf{11}}^{-1}(\theta_1-\mu_1)]/\theta_1,(\sum_{\mathbf{22}}-\sum_{\mathbf{21}}\sum_{\mathbf{11}}^{-1}\sum_{\mathbf{12}})/\theta_1^2\right),\quad(1)$$

where N(**x;a,b**) is the probability density function of a normal vector with mean **a** and variance-covariance matrix **b** evaluated at **x**. One apparent feature of $p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$ is that for large θ_1, the variance-covariance matrix of $(\theta_2, \theta_3)^T$ tends to be small. This is expected as the sizes of the sub-populations get larger, there is more overlap among different sub-populations and their event rates tend to be similar. Another observation is that $(\mu_2, \mu_3)^T$ is the event rate in the entire population for the control and intervention multiplied by p. Thus, roughly speaking, the center of the distribution of treatment effect (e.g. a contrast between θ_2 and θ_3) over sub-populations should be close to the treatment effect in the entire population.

2.3 The Empirical Bayes (EB) Estimation

Let n be the total sample size of the data and r be the probability that a subject is randomized to the intervention arm, both of which are considered fixed throughout this paper. For a given sub-population **Z**, the data can be summarized as **d** = (n_0,n_1, y_0, y_1), where y_0 and n_0 are the count of events and the sample size for the control arm, and similarly y_1 and n_1 for the intervention arm. A natural estimator of $\boldsymbol{\theta} = \boldsymbol{\theta}$(**Z**) is

$$\hat{\boldsymbol{\theta}}=(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3),\ \hat{\theta}_1=(n_0+n_1)/n,\ \hat{\theta}_2=y_0/n_0,\text{and}\ \hat{\theta}_3=y_1/n_1.\quad(2)$$

Throughout this article, we will refer to $\hat{\theta_2}$, $\hat{\theta_3}$ and the odds ratio based on $\hat{\theta_2}$ and $\hat{\theta_3}$ as the *standard estimates*.

The conditional distribution of **d** given $\boldsymbol{\theta}$ can be written as

$$p(\mathbf{d}|\boldsymbol{\theta}) = B(n_0+n_1;n,\theta_1) \times B(n_1;n_0+n_1,r) \times B(y_0;n_0,\theta_2) \times B(y_1;n_1,\theta_3), \quad (3)$$

where $B(x;a,b)$ is the binomial probability mass function evaluated at x with a and b as the number of trials and the success probability. By the Bayes rule, the posterior distribution of $\boldsymbol{\theta}$ is

$$p_{\boldsymbol{\lambda}}(\boldsymbol{\theta}|\mathbf{d}) \propto p_{\boldsymbol{\lambda}}(\boldsymbol{\theta}) \times p(\mathbf{d}|\boldsymbol{\theta}) \propto p_{\boldsymbol{\lambda}}(\boldsymbol{\theta}) \times B(n_0+n_1;n,\theta_1) \times B(y_0;n_0,\theta_2) \times B(y_1;n_1,\theta_3). \quad (4)$$

If $\boldsymbol{\lambda}$ is known, then $p_{\boldsymbol{\lambda}}(\boldsymbol{\theta} \mid \mathbf{d})$ can be used to derive the posterior distribution of (θ_2, θ_3) by integrating out θ_1. The statistical evidence on the treatment effect for the corresponding sub-population can be characterized by the posterior distribution of a contrast between θ_2 and θ_3 (e.g. odds ratio, risk difference). A nice property of this method is the straightforward and intuitively appealing interpretation of the posterior distribution of the treatment effect. If the 5^{th} percentile of the posterior distribution of the odds ratio is 1, then it implies that among ALL sub-populations with the same $\mathbf{d} = (n_0, n_1, y_0, y_1)$, 95% of them (based on weights defined in the prior) will have odds ratios greater than 1. If odds ratio greater than 1 means treatment benefit, then it implies that the treatment is effective in 95% of the sub-populations with the same data as the sub-population of interest. As many of these sub-populations have overlap with the selected sub-population, the stochastic behavior of them has high "relevance".

In practice, $\boldsymbol{\lambda}$ is unknown. A straight forward estimator $\hat{\boldsymbol{\lambda}}$ can be obtained by replacing τ_j, a_j and β_j with sample proportions $\hat{\tau_j}$, $\hat{a_j}$ and $\hat{\beta_j}$ in the definition of $\boldsymbol{\lambda}$. That is, we can obtain $\hat{\tau_j}$ as the sample proportion of the subjects that fall in cell j, and similarly obtain $\hat{a_j}$ and $\hat{\beta_j}$ as the sample event rates within cell j in the control and intervention arms. Then $p_{\hat{\boldsymbol{\lambda}}}(\boldsymbol{\theta} \mid \mathbf{d})$ can be used to make inference about the treatment effect. Since some of the cells are rather small, the estimators $\hat{a_j}$ and $\hat{\beta_j}$ themselves may not be precise. However, they also have small contributions to the variation of $\hat{\boldsymbol{\lambda}}$ due to small values of $\hat{\tau_j}$. On the other hand, larger cells with more precise estimators $\hat{a_j}$ and $\hat{\beta_j}$ will dominate the variation of $\hat{\boldsymbol{\lambda}}$. For cells with no control or intervention units, $\hat{a_j}$ or $\hat{\beta_j}$ can be set to 0.

2.4 Computation of the Posterior Distribution

Parameters of the posterior distribution $p_{\boldsymbol{\lambda}}(\boldsymbol{\theta} \mid \mathbf{d})$ can be computed by standard sampling techniques. Let $\hat{\boldsymbol{\Omega}} = \mathrm{diag}(\hat{\theta_1}(1-\hat{\theta_1}) / n, \hat{\theta_2}(1-\hat{\theta_2}) / n_0, \hat{\theta_3}(1-\hat{\theta_3}) / n_1)$ and $\boldsymbol{\Delta} = \mathrm{diag}(1, 1/\mu_1, 1/\mu_1)$, where the function "diag(**a**)" converts vector **a** into a diagonal matrix. As a binomial distribution can be approximated by a normal distribution,

$$B(n_0+n_1;n,\theta_1) \times B(y_0;n_0,\theta_2) \times B(y_1;n_1,\theta_3) \approx N(\hat{\boldsymbol{\theta}};\boldsymbol{\theta},\hat{\boldsymbol{\Omega}}). \quad (5)$$

We can also approximate $p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$ by a multivariate normal distribution:

$$p_{\boldsymbol{\lambda}}(\boldsymbol{\theta}) \approx N(\boldsymbol{\theta};\boldsymbol{\Delta\mu},\boldsymbol{\Delta}\sum\boldsymbol{\Delta}). \quad (6)$$

Applying approximations of (5) and (6) to (4), we obtain the following importance function:

$$p^*_{\boldsymbol{\lambda}}(\boldsymbol{\theta}|\mathbf{d}) \propto N(\boldsymbol{\theta};\boldsymbol{\Delta}\boldsymbol{\mu}, \boldsymbol{\Delta}\sum\boldsymbol{\Delta}) \times N(\hat{\boldsymbol{\theta}};\boldsymbol{\theta},\hat{\boldsymbol{\Omega}}) \\ =N(\boldsymbol{\theta};\mathbf{U},\mathbf{V}), \quad (7)$$

where $\mathbf{U} = (\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{-1} + \hat{\boldsymbol{\Omega}}^{-1})^{-1}(\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\theta}})$, $\mathbf{V} = (\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{-1} + \hat{\boldsymbol{\Omega}}^{-1})^{-1}$. Since it is easy to sample from $p^*_{\boldsymbol{\lambda}}(\boldsymbol{\theta}|\mathbf{d})$, a large number of samples can be generated to estimate the posterior distribution of $\boldsymbol{\theta}$ with proper weight adjustment. Specifically, let $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)},\ldots, \boldsymbol{\theta}^{(m)}$ be m samples from $p^*_{\boldsymbol{\lambda}}(\boldsymbol{\theta}|\mathbf{d})$, and let $w_i = \frac{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(i)}|\mathbf{d})}{m p^*_{\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(i)}|\mathbf{d})}$ be the weight for $\boldsymbol{\theta}^{(i)}$ (i=1,2,…,m). Then the posterior mean of a function $h(\boldsymbol{\theta})$(i.e. the odds ratio) can be estimated by $\sum_{i=1}^{m} w_i h(\boldsymbol{\theta}^{(i)})$. In addition, the empirical distribution of $h(\boldsymbol{\theta}^{(i)})$ with probability $w_i / \sum_{i=1}^{m} w_i$ can be used to estimate percentiles of the posterior distribution of $h(\boldsymbol{\theta})$. As m goes to infinity, these estimators converge to the true parameters of $p_{\boldsymbol{\lambda}}(\boldsymbol{\theta} \mid \mathbf{d})$. In practice, one can use the m samples to estimate the precision of the estimator in order to choose a proper m. In our analysis in Sections 3 and 4, we set m=100,000.

2.5 Estimation Error

In this particle, we are primarily interested in the posterior percentiles of the odds ratio between θ_2 and θ_3 for fixed $\mathbf{d} = (n_0, n_1, y_0, y_1)$, which can be used to construct empirical Bayes confidence intervals (Carlin and Gelfand, 1990; Rubin, 1984). By the normal-like approximation (1) and the subsequent posterior distribution (4), a percentile can be viewed as a function of $\boldsymbol{\lambda}$, which will be denoted by $F(\boldsymbol{\lambda})$. Let ψ be the true value of the parameter. Then

$$F(\hat{\boldsymbol{\lambda}}) - \psi = [F(\hat{\boldsymbol{\lambda}}) - F(\boldsymbol{\lambda})] + [F(\boldsymbol{\lambda}) - \psi]. \quad (8)$$

Thus there are two sources of error in estimating ψ. The first term on the right side of (8) represents the error due to the estimation of $\boldsymbol{\lambda}$ (sampling variation). If the sample size is N_e, then by the standard maximum likelihood theory $\hat{\boldsymbol{\lambda}}$ is $\sqrt{N_e}$-consistent estimator of $\boldsymbol{\lambda}$ and is asymptotically normal; $F(\hat{\boldsymbol{\lambda}}) - F(\boldsymbol{\lambda})$ is also asymptotically normal by the Delta method with a convergence rate of $N_e^{-1/2}$, as the posterior percentile as a function of $\boldsymbol{\lambda}$ is sufficiently smooth. The second term on the right side of (8) represents the error due to approximating the true prior distribution of $\boldsymbol{\theta}$ by (1). This term contributes to the bias in estimating ψ. As L gets large, we hope that this term tends to be small. For simulation studies and real data analysis in this article, we focus on correcting bias in the first term. In particular, we will use the method proposed by Efron (Efron, 1987) to correct bias in estimates of posterior percentiles that are used to construct the EB confidence limits. Details are provided in the Appendix.

3. APPLICATION TO THE MAGIC DATA

The MAGIC trial (Magnesium in Coronaries (MAGIC) Trial Investigators, 2002) sought to assess the effectiveness of supplemental administration of intravenous magnesium in reducing 30-day all-cause mortality in patients with ST-elevation myocardial infarction (STEMI). The trial was double-blinded with a placebo group as the control arm. A total of 6213 patients were randomized with 3113 and 3100 in the intervention and the control arms, respectively. Within 30 days, 475 (15.3%) and 472 (15.2%) in the intervention and control arms had died (p-value=0.96). Therefore, there was no statistical evidence to support the efficacy of administration of the intravenous magnesium in reducing mortality. Based on the results of MAGIC trial and another study, the 2004 American College of Cardiology/ American Heart Association guidelines on STEMI recommended that routine intravenous magnesium should not be given. Nevertheless, the results of other randomized control trials conducted before MAGIC had led to inconsistent conclusions, with some indicating efficacy and some not (Magnesium in Coronaries (MAGIC) Trial Investigators, 2002). One possible explanation is that the intervention may be helpful for some patients, though not so for others. Nevertheless, no evidence of efficacy was found in the 18 pre-specified sub-populations defined by seven binary and one four-level categorical baseline covariates in the MAGIC study. The eight variables are described in Table 1.

To illustrate the EB method described in Section 2, we focus on three sub-populations. Sub-population A is composed of those with previous myocardial infarction, chest pain at randomization and age ≥65 years without any restriction on the values of other variables in Table 1. Using the notation in Table 1, this sub-population is labelled as “V4=V6=V8=1”. Sub-population B includes those with previous myocardial infarction and age ≥65 years (V4=V8=1) and sub-population C includes those with chest pain at randomization (V6=1). The three sub-populations were chosen as they represent sub-populations with different sizes. Among the 6213 subjects enrolled in the study, 18 had missing values on at least one of the eight variables or the outcome. Thus, our analysis data set is composed of 6195 subjects. The 6195 subjects form 144 cells, each of which has at least one subject. In our analysis, the parameter p for prior was set to be the number of cells included in the sub-population of interest divided by the total number of cells.

Table 2 shows the estimates of the odds ratio of mortality (control over intervention) for the three sub-populations (e.g. odds ratio greater than 1 indicates treatment benefit). The sizes of sub-populations A, B and C are 8.6%, 19.0% and 44.0%. The standard point estimates are fairly similar, ranging from 1.01 to 1.13. Thus, there is little variation in treatment effect in subjects covered by the three sub-populations. To avoid heavy influence of extreme values, we use the posterior median as the point estimate for the empirical Bayes (EB) and bias-corrected empirical Bayes (EB_BC) methods. It is not surprising that the both estimates shrink the standard estimates towards 1 since the odds ratio of mortality of the entire 6195 subjects is essentially 1.00. The standard 95% CIs in Table 2 are frequentists confidence intervals, meaning the confidence intervals as a random quantity covers the true odds ratio 95% of the time under a long run of repeated sampling, On the other hand, the EB 95% CIs essentially are Bayes credible regions, except that they entail frequentists errors in estimating the priors. The EB_CIs have an intuitively appealing interpretation. For example,

the EB 95% CI for sub-population A is 0.77–1.33, which means among all sub-populations with the same data as sub-population A, 95% of them (based on weights defined in the prior) will have a true odds ratio that falls in the region of (0.77, 1.33). The EB_BC 95% CIs try to correct potential bias in the EB CIs so that the confidence intervals on average cover 95% probability mass of the true posterior distribution of the odds ratios under a long run of repeated sampling. It can be seen the EB and EB_BC interval estimates are fairly close. Compared with the standard 95% CI, the EB and EB_BC intervals shrink both the lower and upper limits with a more pronounced effect on the upper limits. Thus, the empirical Bayes CIs trim both the high treatment benefit and high treatment harm ends (particularly the treatment benefit end), because the “experience” of other sub-populations suggests so.

4. SIMULATION STUDIES

We conducted a simulation study to investigate the properties of EB and EB_BC when the estimation process is repeated. In particular, we are interested in whether or not the empirical Bayes confidence intervals indeed cover 95% probability mass with respect to the true posterior distribution when averaged over repeated Monte Carlo samples. We consider the following simulation scheme. Cells with at least one control and one intervention subject from the 144 cells formed by the 6195 subjects of the MAGIC trial are retained. For the retained cells, we assume that the empirical cell size and event rates under control and intervention within each cell are the true population parameters. Thus, in this setting, there is essentially no treatment effect over the entire population (e.g. mortality rates are 15.2% and 15.2% for the control and intervention, respectively). But in 42% of the cells the mortality rate under intervention is lower than control (treatment benefit) and there is treatment harm in the other 58% cells.

We generated 1000 Monte Carlo data sets, each of which is composed of 6195 subjects. For each Monte Carlo data set, we obtained the EB, EB_BC and standard estimates of sub-populations A, B and C. In addition, for each Monte Carlo data set, we computed the true posterior distribution of the odds ratio for the three sub-populations. The EB and EB_BC CIs were then compared to the posterior distribution to calculate the coverage of posterior probability mass. The average of the coverage over the 1000 Monte Carlo data sets is called “mean of probability coverage”. Moreover, we also performed the same estimation tasks for a sub-population D (3 cells, 9.6% of the total population) with stronger treatment benefit based on the population parameters. In Table 3, we provide numerical summary of the simulation results. In terms of point estimate, it is not surprising that the EB and EB_BC pull the point estimate towards 1 due to shrinkage, leading downward bias. This is particularly apparent for sub-population D, where the relative shrinkage is close to 20%. However, the square root of the mean squared error (SRMSE) is always smaller than the standard point estimate, suggesting a benefit in trading bias for precision. The shrinkage of the EB and EB_BC CIs are also reflected by the shift of the 95% CI towards the left, particularly at the upper end, suggesting trimming of strong treatment benefit based on experience of other sub-populations. The EB_CIs, by the way they are constructed, do not necessarily cover the true value with 95% probability under repeated sampling. Thus, no coverage probability in the conventional sense is reported for EB. In addition, due to the estimation of the prior, the EB_CI based on a given Monte Carlo data does not necessarily

cover 95% of the true posterior probability mass. However, as shown in Table 3, the EB_CIs cover about 95% of the true posterior probability mass for sub-populations A, B and D on average (see “Mean of probability coverage of EB 95% CI”), which implies that the coverage probability with respect to the true posterior distribution has little bias for these sub-populations. The EB_CI has a slightly under coverage for sub-population C. This is because the posterior distribution of sub-population C is very sharp due to the relatively large size. Thus, a tiny bias in posterior percentile estimates will translate to relatively large bias in posterior probability coverage. For instance, the average (over the 1000 Monte Carlo data sets) of the 2.5 and 5 percentiles of the posterior distribution of odds ratio for sub-population C is 0.932 and 0.948. Thus, a small bias of around 0.01 in estimating these percentiles (a little over 1% relative bias) will lead to coverage bias of a couple of percentages. EB_BC CI does not offer an obvious improvement over EB_CI, and even tends to have a bit more under coverage for sub-population C.

Overall, the EB method, as expected, shrinks the odds ratio estimate of a sub-population towards the mean of the odds ratios of all sub-populations. The shrinkage represents a different balance in bias-precision trade-off, leading to an improved SRMSE. The relative conservativeness of EB has the advantage of easy control of false positives when a number of sub-populations are being evaluated (Efron, 2010b). In addition, the proposed EB estimation procedure on average covers the nominal probability mass of the true posterior distribution.

5. DISCUSSION

We propose an empirical Bayes method to estimate the treatment effect in a sub-population for a binary outcome. The empirical Bayes offers a natural way to combine both the direct evidence coming from the data of the sub-population of interest and the indirect evidence coming from data of other sub-populations. Our method has three major advantages. First, the posterior distribution of the treatment effect has an appealingly natural and intuitive interpretation. Second, the prior is estimated using data to maintain objectivity. Third, when multiple sub-populations are evaluated at the same time, the posterior distribution offers a straight forward solution for false discovery rate (FDR) estimate (Efron, 2010b).

Although discrete baseline characteristics are considered in this article, the method can be generalized to continuous characteristics by applying appropriate thresholds to discretize them. In particular, as long as the resulting L non-empty cells cover the majority of the population, our approach can be applied. One can apply a sufficient number of thresholds to continuous variables such that the discretized variables still maintain adequate granularity for practical purpose. In fact, a large L has at least three theoretical advantages. First, the asymptotic approximation of equation (1) will work better. Second, we can study the heterogeneity in a finer resolution. Third, if we split cell j into two cells j' and j'', it is clear that

$$\tau_j^2 > \tau_{j'}^2 + \tau_{j''}^2,\ \ \tau_j^2\alpha_j^2 = \tau_j^2\left(\frac{\tau_{j'}\alpha_{j'} + \tau_{j''}\alpha_{j''}}{\tau_j}\right)^2 > \tau_{j'}^2\alpha_{j'}^2 + \tau_{j''}^2\alpha_{j''}^2,$$
$$\tau_j^2\alpha_j\beta_j = (\tau_{j'}\alpha_{j'} + \tau_{j''}\alpha_{j''})(\tau_{j'}\beta_{j'} + \tau_{j''}\beta_{j''}) > \tau_{j'}^2\alpha_{j'}\beta_{j'} + \tau_{j''}^2\alpha_{j''}\beta_{j''}.$$

Therefore, as L gets large, Σ becomes small and provides more prior information. Certainly, as the empirical Bayes method needs to estimate the prior, its performance with different thresholds applied to continuous variables needs to be studied through simulations.

The problem discussed in this paper has some unique features that differentiate it from the setting where Tweedie's formula can be applied (Efron, 2011). The key factor is that our problem has a natural definition of the "prior distribution" without the need to extrapolate beyond the observed units to hypothetically infinite units for the induction of the prior distribution. In this sense, it is conceptually easy and appealing. In addition, because of the overlapping structure of different sub-populations, estimation of the prior can be performed by directly estimating the parameters associated with the prior. Consequently, we can estimate essentially any parameters associated with the posterior distribution. In contrast, Efron's work (Efron, 2011) is based on a general classic empirical Bayes setting. It requires the distribution of the data conditional on the true parameter to follow an exponential family so that the posterior mean and variance can be estimated by Tweedie's formula combined with a model of the marginal distribution of the data. The advantage of Tweedie's formula is that it does not require any information about the prior. The downside is that it is not clear how other posterior parameters such as percentiles can be estimated.

From the methodology perspective, the method described in this article is a proof-of-concept. There are questions that are beyond the scope of this article, which need to be addressed in future studies. First, the accuracy of the estimates of the posterior parameters, particularly the tail percentiles, can be improved. As our simulation studies show, the current percentile estimate still suffers some level of bias when the posterior distribution is sharp. More accurate method to further eliminate bias will greatly enhance the practical utility of this approach. Second, more comprehensive numerical investigations are needed to better understand the performance of this approach under various conditions, such as the sample size, the number of non-empty cells, and the prior parameter λ. Fourth, if a selection process is implemented to select sub-population with strong treatment effect, would the EB method be able to correct potential bias as Efron showed (Efron, 2011)? Answers to these questions will help us better understand the value of the empirical Bayes method in estimating treatment effects in sub-populations.

Acknowledgments

This manuscript was prepared using MAGIC Research Materials obtained from the National Heart, Lung and Blood Institute (NHLBI) Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the MAGIC or the NHLBI.

Funding

This work is supported in part by National Institutes of Health (NIH) grant R21 CA152463 and the Indiana University Health-Indiana University School of Medicine Strategic Research Initiative in Cardiology.

References

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 1995; 57:289–300.

Berger JO, Wang X, Shen L. A Bayesian approach to subgroup identification. J Biopharm Stat. 2014; 24:110–129. [PubMed: 24392981]

Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. Biostatistics. 2011; 12:270–282. [PubMed: 20876663]

Carlin BP, Gelfand AE. Approaches for Empirical Bayes Confidence Intervals. Journal of the American Statistical Association. 1990; 85:105–114.

Davidoff F. Heterogeneity is not always noise: lessons from improvement. The Journal of the American Medical Association. 2009; 302:2580–2586. [PubMed: 20009058]

Davis CE, Leffingwell DP. Empirical Bayes estimates of subgroup effects in clinical trials. Control Clin Trials. 1990; 11:37–42. [PubMed: 2157579]

Dixon DO, Simon R. Bayesian subset analysis. Biometrics. 1991; 47:871–881. [PubMed: 1742443]

Efron B. Empirical Bayes Confidence Intervals Based on Bootstrap Samples: Comment. Journal of the American Statistical Association. 1987; 82:754.

Efron B. Empirical Bayes Methods for Combining Likelihoods. Journal of the American Statistical Association. 1996; 91:538–550.

Efron B. The Future of Indirect Evidence. Statistical Science. 2010a; 25:145–157. [PubMed: 21243111]

Efron, B. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Cambridge, UK: Cambridge University Press; 2010b.

Efron B. Tweedie's Formula and Selection Bias. Journal of the American Statistical Association. 2011; 106:1602–1614. [PubMed: 22505788]

Efron B. A 250-Year Argument: Belief, Behavior, and the Bootstrap. Bulletin of the American Mathematical Society. 2012; 50:129–146.

Foster JC, Taylor JM, Ruberg SJ. Subgroup identification from randomized clinical trial data. Stat Med. 2011; 30:2867–2880. [PubMed: 21815180]

Jones HE, Ohlssen DI, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. Clin Trials. 2011; 8:129–143. [PubMed: 21282293]

Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. The Journal of the American Medical Association. 2007; 298:1209–1212. [PubMed: 17848656]

Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search--a recursive partitioning method for establishing response to treatment in patient subpopulations. Stat Med. 2011; 30:2601–2621. [PubMed: 21786278]

Louis TA. Estimating a population of parameter values using Bayes and Empirical Bayes methods. Journal of the American Statistical Association. 1984; 79:393–398.

Magnesium in Coronaries (MAGIC) Trial Investigators. Early administration of intravenous magnesium to high-risk patients with acute myocardial infarction in the Magnesium in Coronaries (MAGIC) Trial: a randomised controlled trial. Lancet. 2002; 360:1189–1196. [PubMed: 12401244]

Morris CN. Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association. 1983; 78:47–59.

Qian M, Murphy SA. Performance Guarantees for Individualized Treatment Rules. The Annals of Statistics. 2011; 39:1180–1210. [PubMed: 21666835]

Rubin DB. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statisticia. The Annals of Statistics. 1984; 12:1151–1172.

Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. The New England Journal of Medicine. 2007; 357:2189–2194. [PubMed: 18032770]

Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively Selecting a Target Population for a Future Comparative Study. Journal of the American Statistical Association. 2013; 108:527–539. [PubMed: 24058223]

Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. Journal of the American Statistical Association. 2012; 107:1106–1118. [PubMed: 23630406]

APPENDIX

The prior distribution of θ

The elements of **Z** i.i.d. Bernoulli variables with success probability p.

Let $\mathbf{S}(\mathbf{Z}) = (S_1(\mathbf{Z}), S_2(\mathbf{Z}), S_3(\mathbf{Z}))^T$, where $S_1(\boldsymbol{Z})=\sum_{j=1}^{L}\tau_j Z_j, S_2(\boldsymbol{Z})=\sum_{j=1}^{L}\tau_j\alpha_j Z_j, S_3(\boldsymbol{Z})=\sum_{j=1}^{L}\tau_j\beta_j Z_j$. Then $\theta_1(\mathbf{Z}) = S_1(\mathbf{Z})$, $\theta_2(\mathbf{Z}) = S_2(\mathbf{Z})/S_1(\mathbf{Z})$, $\theta_3(\mathbf{Z}) = S_3(\mathbf{Z})/S_1(\mathbf{Z})$.

Suppose as $L\to\infty$, $\mathrm{Max}(\tau_j) = O(L^{-1})$. Given any $\varepsilon>0$, for sufficient large L, $\sqrt{L}\tau_j Z_j \le \sqrt{L}\tau_j<\varepsilon/\sqrt{3}$, and similarly both $\tau_j\alpha_j Z_j$ and $\tau_j\beta_j Z_j$ are bound by $\varepsilon/\sqrt{3}$. Then $\|\sqrt{L}(\tau_j Z_j, \tau_j\alpha_j Z_j, \tau_j\beta_j Z_j)\|<\varepsilon$ for all j. By Linderberg-Feller central limit theorem, as $L\to\infty$,

$$\sqrt{L}\,(\mathbf{S}-E(\mathbf{S})) \to N\left(0, \lim_{L\to\infty}(L\times Var(\mathbf{S}))\right).$$

Therefore, $\mathbf{S}(\mathbf{Z}) \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu}=(\mu_1,\mu_2,\mu_3)^T=E(\mathbf{S})=p\left(1,\sum_{j=1}^{L}\tau_j\alpha_j,\sum_{j=1}^{L}\tau_j\beta_j\right)^T$$

$$\sum=Var(\mathbf{S})=\begin{pmatrix}\sum_{\mathbf{11}} & \sum_{\mathbf{12}}\\ \sum_{\mathbf{21}} & \sum_{\mathbf{22}}\end{pmatrix},\ \sum_{\mathbf{11}}=p(1-p)\sum_{j=1}^{L}\tau_j^2, \sum_{\mathbf{12}}=\sum_{\mathbf{21}}^{T}=p(1-p)\left(\sum_{j=1}^{L}\tau_j^2\alpha_j,\sum_{j=1}^{L}\tau_j^2\beta_j\right),$$

$$\sum_{\mathbf{22}}=p(1-p)\begin{pmatrix}\sum_{j=1}^{L}\tau_j^2\alpha_j^2 & \sum_{j=1}^{L}\tau_j^2\alpha_j\beta_j\\ \sum_{j=1}^{L}\tau_j^2\alpha_j\beta_j & \sum_{j=1}^{L}\tau_j^2\beta_j^2\end{pmatrix}.$$

Then $(S_2, S_3)^T \mid S_1 \sim N(\mu^*(S_1), \Sigma^*)$, where $\boldsymbol{\mu}^*(S_1)=(\mu_2,\mu_3)^T+\sum_{\mathbf{21}}\sum_{\mathbf{11}}^{-1}(S_1-\mu_1), \sum^*=\sum_{\mathbf{22}}-\sum_{\mathbf{21}}\sum_{\mathbf{11}}^{-1}\sum_{\mathbf{12}}$. Let $\boldsymbol{\lambda} = (\boldsymbol{\mu},\boldsymbol{\Sigma})$. It follows that the distribution of $\theta = \theta(\mathbf{Z})$ can be written as

$$p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})=p_{\boldsymbol{\lambda}}(\theta_1)p_{\boldsymbol{\lambda}}(\theta_2,\theta_3|\theta_1)=N(\theta_1;\mu_1,\sum_{\mathbf{11}})\times N(\theta_2,\theta_3;\boldsymbol{\mu}^*(\theta_1)/\theta_1,\sum^*/\theta_1^2)$$
$$=N(\theta_1;\mu_1,\sum_{\mathbf{11}})\times N\left((\theta_2,\theta_3)^T;(\mu_2,\mu_3)^T+\sum_{\mathbf{21}}\sum_{\mathbf{11}}^{-1}(\theta_1-\mu_1)/\theta_1,(\sum_{\mathbf{22}}-\sum_{\mathbf{21}}\sum_{\mathbf{11}}^{-1}\sum_{\mathbf{12}})/\theta_1^2\right)$$

The bootstrap procedure to correct bias in posterior percentile estimates

Let $\eta = \eta(\theta)$ be the odds ratio and $\phi(\tau_j, \alpha_j, \beta_j, j = 1, 2, \ldots, L)$. Denote by $C(\mathbf{d}, \hat{\lambda}, \alpha)$ the $\alpha \times 100\%$ percentile of the estimated posterior distribution of η. The idea is to find an α' such that $E(p_{\lambda}(\eta \leq C(\mathbf{d}, \hat{\lambda}, \alpha') \mid \mathbf{d}) = \alpha$, where the expectation is with respect to the true sampling distribution of the entire data given the true parameter ϕ and sub-population data $\mathbf{d}$, $p(.|\phi, \mathbf{d})$ (Carlin and Gelfand, 1990; Rubin, 1984). In other words, the sample space of $p(.|\phi, \mathbf{d})$ includes all data sets that yield the same data $\mathbf{d}$ for the sub-population of interest. We can then correct $C(\mathbf{d}, \hat{\lambda}, \alpha)$ by $C(\mathbf{d}, \hat{\lambda}, \alpha')$ so that the posterior percentile estimates yield the nominal percentile coverage on average. We cannot directly solve α' as we do not know ϕ. A bootstrap method can be used to estimate α' by solving

$$\frac{1}{N_b}\sum_{i=1}^{N_b} p_{\hat{\lambda}}(\eta \leq C(\mathbf{d}, \hat{\lambda}_{\mathbf{i}}^{*}, \alpha')|\mathbf{d}) = \alpha, \quad (9)$$

where $\hat{\lambda}_{\mathbf{i}}^{*}$ is the estimate from the ith bootstrap sample that was generated from $p(.|\phi, \hat{\mathbf{d}})$ and N_b is the number of bootstrap samples.

To generate a bootstrap sample from $p(.|\phi, \hat{\mathbf{d}})$, we need to condition on data $\mathbf{d}$. Let S_1 be the set of $n_0 + n_1$ subjects in the sub-population of interest and S_2 be the set of n_r subjects that do not belong to the sub-population of interest. We first draw n_r subjects with replacement from S_2 just like the standard bootstrap method. Then we draw $n_0 + n_1$ subjects from S_1, while maintaining the total number of control, intervention, events under control and events under intervention at n_0, n_1, y_0, and y_1, respectively. This can be easily done using multinomial distributions. Let C be the list of cells in the sub-population of interest and $\mathbf{d}^{(j)} = (n_0^{(j)}, n_1^{(j)}, y_0^{(j)}, y_1^{(j)})$ be the data for cell $j \in C$. Let

$$w_0^{(j)} = (n_0^{(j)} + n_1^{(j)}) y_0^{(j)} / n_0^{(j)}, v_0^{(j)} = (n_0^{(j)} + n_1^{(j)})(n_0^{(j)} - y_0^{(j)}) / n_0^{(j)}$$
$$w_1^{(j)} = (n_0^{(j)} + n_1^{(j)}) y_1^{(j)} / n_1^{(j)}, v_1^{(j)} = (n_0^{(j)} + n_1^{(j)})(n_1^{(j)} - y_1^{(j)}) / n_1^{(j)}.$$

Then the number events under control in cells of C can be drawn from a multinomial distribution $\mathrm{Multi}(y_0, w_0^{(j)} / \sum_{j \in C} w_0^{(j)})$, where the size is y_0 and probability vector is $w_0^{(j)} / \sum_{j \in C} w_0^{(j)}$. Similarly, the number of non-events under control, the number events under intervention, and the number of non-events under intervention can be drawn from $\mathrm{Multi}(n_0 - y_0, v_0^{(j)} / \sum_{j \in C} v_0^{(j)}), \mathrm{Multi}(y_1, w_1^{(j)} / \sum_{j \in C} w_1^{(j)})$ and $\mathrm{Multi}(n_1 - y_1, v_1^{(j)} / \sum_{j \in C} v_1^{(j)})$.

Table 1

The eight baseline covariates used to define the 18 pre-specified sub-populations in the original MAGIC publication.

Variable	Value
Stratum (V1)	1: candidates for reperfusion therapy, 2: otherwise
Time from myocardial infarction to bolus (hour) (V2)	1: ≤1, 2: 1–3, 3: 3–6, 4:>6
History of diabetes (V3)	0: no, 1: yes
Previous myocardial infarction (V4)	0: no, 1: yes
Received reperfusion (V5)	0: no, 1: yes
Chest pain at randomization (V6)	0: no, 1: yes
Type of reperfusion (V7)	0: No reperfusion attempt; 1: lytics, 2: Percutaneous transluminal coronary angioplasty (PTCA)
Age (years) (V8)	0: <65, 1: ≥65

Table 2

Estimation of the odds ratio (control over treatment) of 3-day mortality for three sub-populations. EB_BC is based on 500 bootstrap samples. EB: empirical Bayes; EB_BC: empirical Bayes with bias correction

	Sub-population A (V4=V6=V8=1)	Sub-population B (V4=V8=1)	Sub-population C (V6=1)
# of cells	26	48	78
Size	8.6%	19.0%	44.0%
Standard point estimate	1.13	1.10	1.01
Standard 95% CI	0.73–1.73	0.82–1.47	0.82–1.23
EB point estimate *	1.03	1.04	1.01
EB 95% CI	0.77–1.33	0.87–1.20	0.91–1.13
EB_BC point estimate	1.00	1.04	1.01
EB_BC 95% CI	0.77–1.23	0.87–1.20	0.92–1.14

* Median of posterior distribution

Table 3

Estimates of the odds ratio (control vs. intervention) of mortality for the three sub-populations based on 1000 Monte Carlo simulations. EB_BC is based on 500 bootstrap samples. EB: empirical Bayes; EB_BC: empirical Bayes with bias correction; SRMSE: square root of mean squared error.

	Sub-population A (V4=V6=V8=1)	**Sub-population B (V4=V8=1)**	**Sub-population C (V6=1)**	**Sub-population D (3 cells, size=9.6%)**
True value	1.29	1.14	1.09	1.53
Mean of standard point estimate (SRMSE)	1.32 (0.29)	1.15 (0.18)	1.10 (0.11)	1.59 (0.42)
Mean of standard 95% CI[#]	0.85–2.05	0.86–1.54	0.90–1.35	0.95–2.66
Mean of EB point estimate[*] (SRMSE)	1.20 (0.23)	1.08 (0.14)	1.05 (0.10)	1.24 (0.33)
Mean of EB 95% CI	0.84–1.68	0.86–1.34	0.92–1.21	0.91–1.72
Mean of probability coverage of EB 95% CI[+]	95.6%	94.6%	92.0%	95.7%
Mean of EB_BC point estimate (SRMSE)	1.17 (0.23)	1.07 (0.15)	1.04 (0.09)	1.21 (0.35)
Mean of EB_BC 95% CI	0.83–1.62	0.86–1.31	0.92–1.19	0.91–1.64
Mean of probability coverage of EB_BC 95% CI	95.1%	93.7%	90.8%	95.2%

[#]Average of the lower and upper limits over the 1000 Monte Carlo samples

[*]Average of the median of posterior distribution over the 1000 Monte Carlo samples

[+]Average of the coverage probability (with respect to the true posterior distribution) over 1000 Monte Carlo samples